PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [6] : G06F 17/30	A1	(11) International Publication Number: **WO 99/13414** (43) International Publication Date: 18 March 1999 (18.03.99)

(21) International Application Number: PCT/GB98/02636

(22) International Filing Date: 2 September 1998 (02.09.98)

(30) Priority Data:
9718905.4   5 September 1997 (05.09.97)   GB

(71) Applicant (*for all designated States except US*): CAMBRIDGE CONSULTANTS LIMITED [GB/GB]; Science Park, Milton Road, Cambridge CB4 4DW (GB).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): MARTIN, Sean, Christopher [GB/GB]; 20 Clare Street, Cambridge CB4 3BY (GB). SHARP, David, William, Nathaniel [GB/GB]; 58a Natal Road, Cambridge CB1 3NY (GB).

(74) Agents: BERESFORD, Keith, Denis, Lewis et al.; Beresford & Co., 2-5 Warwick Court, High Holborn, London WC1R 5DJ (GB).

(81) Designated States: JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: DATA STORAGE AND RETRIEVAL
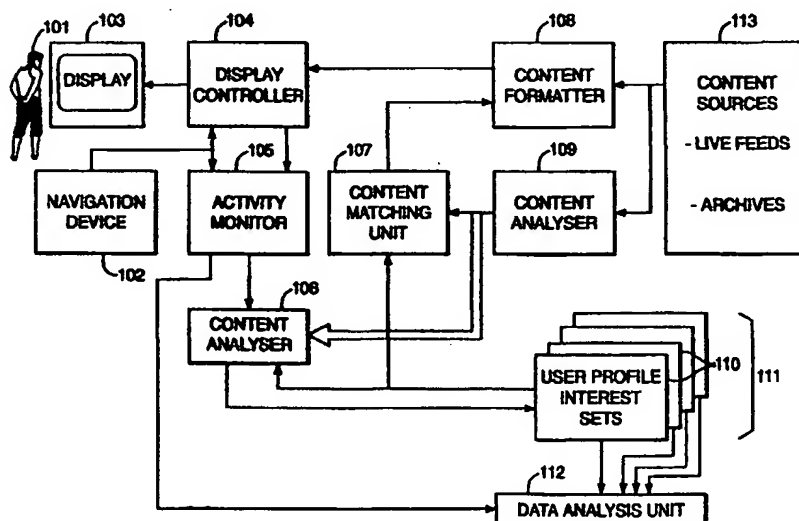
(57) Abstract

There is described a data selection and retrieval system wherein items of information can be retrieved by a user from a database. Control signals input by the user when manipulating retrieved data items are processed to indicate the level of interest which the user has in the data item. The user's level of interest and the content of the displayed data item are then compiled to form a list of the subject–matter in which the user has evinced an interest, and data items in the storage means which relate to similar subject–matter are then indicated as being of likely interest to the user.

1

DATA STORAGE AND RETRIEVAL

The present invention relates to data selection and retrieval, and is particularly concerned with directing a user of an information source or a database towards
5   data items which may be of interest to that user.

Information retrieval systems are known wherein a user selects, for example, news items from a list of headlines, and then retrieves either a synopsis of the news article or the full text for further reading. In
10 order to make such retrieval systems more efficient, it is known to request a user to indicate topics in which he has an interest, so that a listing of headlines may be arranged to present those topics of interest at the head of the list, or even in a separate "personalised"
15 list individual to that user. To achieve this, the user conventionally indicates areas or topics of interest by entering a series of key words to form a search list, and the items in the information source or database are then compared with the search list words to allocate an
20 "interest factor" to the items in relation to that search list. Items with high "interest factors" are presented to the user either as a personalised list, or as highlighted items in a comprehensive list, or as items appearing at the head of a list of available data items.

25   In practice, however, it is found that users are unskilled in the compilation of search lists and often omit key words, or over specify to a degree of detail

2

which hinders the identification of all interesting data items. Furthermore, the user is burdened with the task of editing the search list whenever a new topic is to be added or a topic is to be deleted from the list, or

5   indeed when the user's interest changes from one aspect of a particular topic to another. The editing burden placed on the user, and the necessary skill in initially compiling the key word list, combine to reduce the attractiveness and effectiveness of such a data retrieval

10   system to the user.

It is a concern of the present invention to provide a data retrieval system and method which can discriminate between data items of interest to a user and other data items without the need for the user actively to input

15   instructions relating to his interests.

The present invention is further concerned with a method and apparatus for identifying data items of interest to a user from a plurality of data items stored in a memory, based on monitoring control signals input

20   by the user during selection and display of a data item.

The present invention may further provide a filtering tool which can be incorporated into a conventional browser in order to give a user an initial indication of the likelihood of his being interested in

25   a particular item of information, by analysing the information item to assign content descriptors to it, and then comparing the content descriptors with content

descriptors in a file of a user's personal interests.

A principal feature of the invention is the compiling of a list of the subjects which are of interest to the user by observing the user's reaction to displayed data items. In preferred embodiments the user's reaction is observed by monitoring the user's control of screen attributes such as scrolling speed, or by monitoring the time interval during which a data item is displayed, and inferring the user's interest level from such monitoring measurements by assuming, for example, that scrolling speeds within a certain range indicate careful reading of the item and thus denote high interest, and that higher scrolling speeds indicate that the user is merely 'skimming through' the data item, and has a lower level of interest in its content.

The present invention is capable of inferring not only a user's level of interest in an entire data item, but also interest in specific sections of a data item. For instance a user may skim through the first 20 pages of a document, then carefully read two or three paragraphs, before skimming through the rest of the document - stopping to look carefully at the diagram on page 35 - and then moving on to another data item; the present invention, unlike any prior art systems is capable of detecting and using such information regarding the user's interests.

The personal interest data is compiled by analysing

4

the user's response to items of information presented, by correlating the display control inputs made by the user while the item is being displayed with the content descriptors of the data and inferring from the control

5   inputs the level of interest of the user in the information presented, and incorporating into the user's personal interest data any content descriptors of data items wherein the level of interest exceeds a threshold.

In accordance with a first aspect, a data selection

10 and retrieval system comprises a memory wherein a plurality of data items are stored, first content analysis means for associating one or more content descriptors with each of the respective data items, a display and control means associated with the display to

15 select a data item for display to a user, monitoring means correlating the momentary content of the display with the control signals input by the user to infer an interest level of the user in relation to the momentarily displayed data, user profiling means to assemble a record

20 of the content descriptors relating to data items inferred by the monitoring means to have been of interest to the user, and matching means to compare the content descriptors of data items in the memory with the content descriptors recorded in the user profiling means, the

25 data items whose content descriptors match those stored in the user profiling means being indicated preferentially on the display.

5

In a second aspect, a data selection and retrieval apparatus comprising a first memory means for storing a plurality of data items and associated content descriptor information:

5   selection and display means operable by control signals input by a user to select a data item from the first memory for display, and to manipulate the displayed data item;

monitoring means to receive and monitor the control

10   signals input by the user;

inferring means to infer a level of interest on the part of the user in a data item displayed on the display means, based on the control signals input while the data item is displayed;

15   user profile recording means to record a correlation between the content descriptor of a data item and the level of interest inferred in relation thereto;

determining means to determine as being "of interest" those content descriptors in respect of which

20   the level of interest exceeds a determined threshold;

matching means to compare the content descriptors of the data items in the first memory with the content descriptors determined to be "of interest"; and

listing means to list the data items whose content

25   description are "of interest".

A third aspect concerns a method of identifying data items of interest to a user from a plurality of data

6

items stored in a memory with associated content
descriptor information, wherein a user may input control
signals to a selection and display means so as to select
and display data items, characterised in that the control
5   signals input by the user are monitored and correlations
between control signals and display content are made, and
a level of interest of the user in relation to displayed
data is inferred from that correlation. The method is
applicable not only in data retrieval operations from
10   such sources as news databases or in 'electronic
shopping', where users select and order merchandise from
an electronically stored catalogue, but also in
situations in which it is desired to capture user
reaction data in respect of displayed material without
15   the user having to take specific action to provide
relevance feed-back. For example, the proprietor of a
web site may wish to monitor the reactions of visitors
to the contents of files served from his web site, so as
to gauge the effectiveness of the site as an advertising
20   tool. Merely knowing that a file has been downloaded
does not necessarily give any indication as to the level
of attention with which the file was perused.

A fourth aspect concerns a method of selecting data
items likely to be of interest to a user from a number
25   of data items stored in a memory, wherein the content of
the data items is analysed and one or more content
descriptors is associated with each data item and wherein

7

a user inputs control signals to a selection and display
device to select and display a data item and a monitoring
means also receives the control signals and determines
from them the content descriptor or descriptors of the
5   data item being displayed and also infers a level of
interest in the displayed data item based on the control
signals, a user profile recording means recording the
content descriptors of data items inferred to have been
of interest to the user, and matching means comparing the
10   content descriptors stored in the user profile recording
means with the content descriptors of the data items in
the memory to generate a listing of those data items
having content descriptors corresponding to those in the
user profile recording means.

15   Embodiments of the present invention will now be
described with reference to the accompanying drawings,
in which:

        Figure 1 is a block diagram of a first data
selection and retrieval system;

20   Figure 2 is a block diagram of the server station
of the system of Figure 1;

        Figure 3 is a block diagram of a user station of the
system of Figure 1.

        Figure 4 is a flow chart showing the operation of
25   the system of figures 1 to 3.

        Figure 5 is a diagram showing data stored in the
look-up table 24a of Figure 2.

8

Figure 6 shows the principal elements of the logical architecture of the system.

Referring to Figure 6, the display 103 presents visual, audio and, if appropriate, other sensory
5   information to the user 101. The display 103 may be an LCD, TV, monitor or other visual display unit and is controlled by the display controller 104.

The user 101 receives audio-visual information from the display 103 and reacts to it by causing the
10   navigation device 102 to send out signals to the display controller 104 and activity monitor 105. The navigation device 102 may be a two-dimensional pointing device such as a joystick or mouse or other user input device.

The display controller 104 updates the display 103
15   in response to the user's signals applied to the navigation device 102. The display controller 104 obtains the content to display from the content formatter 108.

The content formatter 108 obtains content from one
20   or more content sources 113. A content source may be a live feed or an archive or other source and may contain information such as the pages of an electronic newspaper, on-line share price information, the pages of an electronic catalogue or other information. The content
25   may be segmented into a number of "articles" - for example an article in a newspaper or an item in an electronic catalogue, and may comprise a still picture,

9

a moving picture, an audio source, or a combination of text and/or pictures and/or audio.

The content formatter 108 summarises and arranges the articles to simplify user navigation through the content material available. For example the content formatter 108 may produce a high-level summary of each article (e.g. just its title) so that the titles of several articles can be viewed on the display 103 simultaneously. It may also produce an intermediate summary of each article, for example consisting of the headline, a couple of lines of text of the article and a picture, or consisting of main sub-headings or paragraph headings if available. The content formatter 108 can also supply the full detail or text of each article on request.

The display controller 104 presents information from the content formatter 108 to the display 103. The display controller 104 can present the high-level summaries of several articles on the screen, and the navigation device 102 may be used to scroll the display up or down to reveal further article summaries in response to the user's input to the navigation device 102. The display controller 104 can also display a higher level of detail about an article or a group of articles, for example in the form of intermediate summaries of articles, in response to a user's request via the navigation device 102. The display controller

10

104 can also show the full detail or text of each article in response to appropriate control signals from the user 101 via the navigation device 102. The display controller 104 can also scroll the full detail or text

5   up and down to reveal more of the full detail or text at the user's request.

The activity monitor 105 gathers information about the material that the user 101 has caused to be displayed. The activity monitor 105 may determine:

10 • which parts of the material available were not selected by the user 101 for display.

• which parts of the material were requested for display by the user 101 and which level of detail the user 101 requested.

15 • the length of time each part of the material that was requested was on the display 103.

• the speed at which the user was scrolling the screen display 103 in response to the material on it.

20 From this information the activity monitor 105 may estimate which parts of the material the user 101 read in detail, which parts were skim-read, which parts were only looked at briefly, and which parts were ignored completely. The activity monitor 105 may attach

25 weighting information to the material or parts of the material (e.g. words, pictures, paragraphs, articles or other sub-divisions) that indicate the estimated level

11

of interest of the user 101 in that part of the material. This weighting information about the material is fed into the content analyser 106.

An important feature is that this weighting
5   information is determined without the user being required to give any explicit indications or relevance feedback - the user was just browsing/reading the information. The fact that the user was viewing the displayed material at reading speed, for example by scrolling the text down
10   the screen, is used to deduce automatically that the user was probably reading the material.

The content analyser 106 takes in the weighted material from the activity monitor 105 and updates the user profile 110 of the user 101 to reflect the level of
15   interest in material that has been read. The level of interest may be increased in response to reading an article in the same area as an existing area of interest. The level of interest may be decreased in response to not reading an available article in an existing area of
20   interest. The user may additionally be given opportunities to provide explicit information in relation to his interests, such as by editing his own user profile.

The content analyser 106 analyses the content of an
25   article or part of an article in which the level of interest of the user 101 has been inferred or where the user 101 has explicitly indicated the level of interest.

12

Where the content of an article or part of an article has already been analysed by content analyser 109 this analysis may be reused by content analyser 106.

The weighting information and the content analysis information is used to update the existing user profile interest sets 110 and to create new interest sets. The method of updating the user profile interest sets 110 may take a number of forms including the following: the level of interest may be increased in response to reading an article in the same area as an existing area of interest. The level of interest may be decreased in response to not reading an available article in an existing area of interest. A new area of interest is created where an article or part of an article is inferred or explicitly indicated by the user 101 to be of interest to the user 101, but there is no existing corresponding interest set in the user profile interest sets 110.

The user profile 110 may consist of weighted sets of keywords or other structured information that indicates the level of interest of the user in various content areas. The user profile 110 may be updated over time so that passing interests are removed over a period of time.

The profiles of many users 111 can be kept and each user profile 110 handled separately.

The profile of a user 110 or many users 111 may be analysed by a data analysis unit 112 for extracting

13

information about what is of interest either to particular users or groups of users, or about the general interest levels of users in relation to categories of data. For example this information could indicate the

5   level of interest shown in particular items in a catalogue, or the parts of a newspaper that were most popular. This information could be used to target advertising material, to give special offers to users who had browsed an item in a catalogue many times without

10   purchasing, or could be used for other purposes.

The user profile 110 may be used to personalise the content of the newspaper, catalogue or other material presented to the user 101 as follows. A content analyser 109 creates a weighted list of keywords (or other content

15   descriptor) characterising pieces of the content from the content sources 113. The content matching unit 107 uses the content descriptor from the content analyser 109 and the user profile 110 to deduce which pieces of content are of greater interest to the user 101. A listing of

20   these interesting pieces of content is prepared for presentation to the user 101 by the content formatter 108. The content formatter 108 may create a "personalised news" section of a newspaper in which to include the interesting pieces of content. Alternatively

25   some other personalised information presentation format may be used in place of, or in addition to, the existing content format. The content formatter 108 may also

14

generate an alert which is communicated to the user through the display 103 or via other communication mechanisms, for example by telephone, messaging or email.

The data analysis unit 112 may also interpret
5   information from the activity monitor 105 for a variety of purposes including evaluation of the effectiveness of the user interface, compiling statistics about the access patterns to articles, portions of articles or groups of articles, and determining the effectiveness of the
10 information in a catalogue.

The system components could be co-located in a single computer, or they may be accommodated by a number of computers and communicate with each other over a network. The only physical constraint is that the
15 display 103 and navigation device 102 must be simultaneously available to the user 101.

An embodiment of the system will now be described. The information retrieval and display system of Figure 1 comprises a server station 10 linked via a distribution
20 system 20 to a plurality of user stations 30. The distribution system 20 may be a local or wide area network, a public or private telecommunication network, the Internet, or any other suitable transmission means providing for two-way traffic between each user station
25 30 and the server 10. Each user station 30 is operable to request data to be downloaded to it from the server 10 by transmitting signals to the server 10 via the

15

distribution network 20. In response to such requests, server 10 retrieves the data from memory, addresses it and sends it over the distribution network 20 to the user station 30. The user may then display the information

5   and peruse it either in detail or briefly, or the user may simply inspect the data and discard it. The user station 30 sends to the server 10 information relating to the control signals input by the user to navigate within a data item displayed at the user station. The

10   user may request further data from the server 10 when he has finished browsing the previous data.

Figure 2 is a block diagram showing the server station 10. The server station 10 comprises a mass storage or memory 21 wherein data items D1, D2, D3 are

15   stored. Each data item has associated with it a content descriptor CD1, CD2, CD3, which may be a sequence of key words extracted from the content of D1, or may represent the content of D1 in some other fashion.

Content descriptors CD1, 2, 3... are assigned to

20   data items D1, D2... by a content analysis unit 22. The content analysis unit may use conventional techniques applying statistical algorithms to the text content of the data item to produce a list of weighted key words which are associated with the data item as its content

25   descriptor. Alternative methods may be used, such as techniques including neural networks, or manual classification of data item contents by an operator

16

reading the item and allocating key words as its content descriptor. The content analysis unit is controlled by the computer 23, so as to ensure that every data item D1 ... in the memory 21 has associated with it a content

5   descriptor CD1...etc.

Server station 10 further includes a user profile generator 24 whose function is to compile an assimilation or record of the content descriptors of data items in which a user has shown interest. The user profile

10   generator 24 in this embodiment includes a look-up table 24a, which stores data correlating the number of words of text displayed on the user's screen with corresponding first and second ranges of screen scrolling speeds which indicate careful reading or speed reading of the screen

15   contents, respectively. The user profiles of a plurality of users ABC...etc are stored in a user profile register 25.

The server 10 may optionally further include a user profile data analysis unit 26 which can access the user

20   profile data register 25 to compile statistical data, for example, in the form of lists of users having similar interests, users having related patterns of access, or users in the same geographical area.

A matching unit 27 operates to compare the content

25   descriptors of data items in the memory 21 with the content descriptors listed in a user's profile, so as to determine which data items relate to information likely

17

to be of interest to that user.

The content analysis unit 22, profile generator 24 and matching unit 27 are preferably implemented as software modules stored in a memory separate from the mass storage memory 21 but accessible by the processor 23. The user profile data register 25 is likewise preferably implemented as a software module, as is the data analysis unit 26, again preferably located in a memory separate from the main storage memory 21 an accessible to processor 23.

Figure 3 is a detailed block diagram of a user station 30, which comprises a conventional display device 31, for example a liquid crystal display or a cathode ray tube. An interface unit 32 is connected between the distribution network 20 and the display 31 and a control device 33 is connected to the interface unit 32. The control device may be a keyboard, a mouse, or a joystick device, depending on the user's preference, and enables a user to direct commands to the server station 10 and/or the display 31.

The interface unit 32 comprises transmitting and receiving apparatus 34 connected to the distribution system 20 for transmitting to the server station 10 requests for downloading of data and for receiving the requested data therefrom, audio and video output circuitry 35 for supplying audio and/or video signals, for example in PAL, NTSC or SECAM form to a television

18

receiver acting as the display 31, and a decoding and
encoding arrangement 36 for decoding signals received
from the transmitter/receiver apparatus 34 and for
encoding signals for supply to the transmitter/receiver
5 apparatus 34 for requesting information from the server
10 and for encoding signals into appropriate form for
supply to the audio and video output circuitry 35. In
addition, a central processor unit 37 is connected to the
transmit/receive apparatus 34, decoder/encoder
10 arrangement 36 and audio and video output circuitry 35
for controlling the operation thereof in accordance with
programs stored in a ROM 38. RAM 39 is provided in the
interface unit 32 and connected to the CPU 37 so that the
CPU 37 may store in the RAM 39 data downloaded from the
15 server station 10 and may retrieve such data from the RAM
39 for appropriate encoding for output as video and/or
audio signals to the display 31.

As shown in Figure 3, the controller 33 is connected
to the CPU. The ROM 38 contains programs for causing the
20 interface unit 32 to respond to movements of the control
device 33 (schematically shown as a joystick in the
Figure) for facilitating browsing of the information
available from the server 10, and to enable control
signals from the control device 33 to be relayed to the
25 server 10 for processing. In the preferred embodiment,
the control device 33 is a joystick and the ROM 38
contains programs which cause the screen display to be

19

scrolled up and down in response to upward and downward
deflections of the joystick, with the scrolling speed
being proportional to the amount of deflection of the
joystick away from a central position. The control

5   device 33 may, however, be any suitable control device,
such as, for example, a mouse, a rocker switch, a
plurality of switches, a wand, a trackball, a touch
screen etc.

The interface unit 32 may be of conventional

10   construction and arrangement and thus may comprise, for
example, a conventional so-called "set-top box" for
connection to a television receiver but containing novel
control programs in ROM 38. In any of the above
situations, ROM such as ROM 38 may be replaced by RAM,

15   in which case the control programs may be transferred to
the RAM via a storage device, for example a conventional
computer disk, or may be transmitted as signals thereto,
for example via the Internet. The control programs could
be transmitted to the user station 30 from the server

20   station 10, for example as Java applets or in other
formats.

The operation of the system illustrated in Figures
1 to 3 by a user to retrieve data will now be explained,
with reference to the flow chart of Figure 4. In the

25   memory 21 of the server 10, a number of data items D1,
D2, D3...etc are stored in association with respective
content descriptors CD1, CD2, CD3. These data items may

20

be input into the memory 21 via an input device associated with the processor 23, such as a keyboard, disk drive, or scanner. Alternatively the data items D1, D2 etc may be received via the distribution network 20.

5   Data items may be input into the memory 21 either with or without their associated content descriptors CD1, CD2 etc. The processor 23 includes means to determine whether each incoming data item has a content descriptor, and may cause the content analysis unit 22 in the server

10   10 to analyse the text of the data item and allocate content descriptors CD1 etc to those data items having no descriptors as each item is input. Alternatively, if a number of data items are received in a batch, for example via a disk drive or the distribution network 20,

15   the processor 23 may determine which data items have no descriptors, and may cause analysis unit 22 to analyse those data items and allocate content descriptors after all items in a batch have been stored.

In a further alternative, in cases where data items

20   in the memory 21 are edited and updated periodically, the processor 23 may cause analysis unit 22 to re-analyse data items after they have been edited and re-allocate content descriptors to edited data items. Re-analysis may be done immediately after editing of a data item, or

25   may be done at predetermined intervals so that every data item added or edited since the last re-analysis is allocated updated content descriptors.

21

The data retrieval process of Figure 4 commences
with a 'log on' step S1, wherein a user requests access
to the data in memory 21, and inputs a user
identification code (User ID) which serves to identify
5　the user for billing and other purposes. The user ID
may, for example, be furnished by a smart card, by using
a terminal at a designated address on the network, or by
using a terminal designated as that user's unique
terminal.

10　　　When a request is received, the processor 23
proceeds to step S2 and interrogates the memory 21 to
generate an index list of all the data items in the
memory, briefly indicating their nature. This may for
example be a listing of the headlines in an electronic
15　newspaper.

Processor 23 then, in step S3, compares the user ID
with the register of user profiles 25 to determine
whether a user profile already exists for this user. If
no user profile exists, then the entire index list is
20　sent to the user's display (step S4) so that the user may
select an item for detailed study. If a user profile
already exists for this user ID in the user profile
register 25, the processor proceeds to execute a routine
to compare the content descriptors CD1, CD2 etc of data
25　items D1, D2 etc with the content descriptors in the user
profile. During this routine, the processor assembles
a list of the data items considered to be 'of interest'

22

to the user. This routine comprises steps S5 to S11.

In steps S5 and S6, data flags are cleared and the user profile of content descriptors CD known to be of interest to the user is obtained. In step S7, a data

5   item has its content descriptor CD compared with the content descriptors in the user profile, and if the descriptor CD matches the descriptors listed in the profile, then the process proceeds to step S8 wherein that data item D is added to an 'of interest' list. The

10   process then proceeds to step S9 where a flag is attached to the data item D. ˙If in step S7 the descriptor CD of the data item D does not match the content descriptors in the user profile, then the process flows directly to step S9 and a flag is attached to the data item D without

15   adding the data item to the 'of interest' list.

In step S10 it is determined whether any unflagged data items remain, and if they do the process passes to step S11 to select the next data item and then to step S7 to compare the content descriptors of the new data

20   item with those in the user profile.

If no unflagged data items remain in step S10, it is determined in step S12 whether an 'of interest' list exists. If not, the process flows to step S4 to display the index list to the user. If an 'of interest' list

25   exists, the process flows to step S13 and both the index list and the 'of interest' list are sent to the user for display.

23

The form of the display of both lists may for example be by displaying the index list and highlighting the items of the index list that are also in the 'of interest' list. Alternatively, the index list may be displayed so that the items also on the 'of interest' list appear at the top of the index list. In a third alternative, the items on the 'of interest' list may be displayed as a separate 'personal' list in addition to or optionally instead of the index list.

The processor 23 then determines, in step S14, whether the user has made a selection from the listed data items for more detailed study. If an item has been selected, then in step S15 the data item is displayed on the user's display 31. By operating the control device 33, the user can generate data item selection signals which are sent to the processor 23 and the processor 23 responds by retrieving the selected data item from memory 21 and sending it to the RAM 39 at the user station 30. The CPU 37 at the user station then causes the data item to be displayed on the screen 31, and data signals indicating the content of the screen are also sent to the processor 23 at the server 10. Screen control signals input from the control device 33 to the CPU 39 enable the user to scroll the displayed data, and navigate between pages of the data item. The user's screen control signals from the control device 33 are also sent to the processor 23 at the server station 10.

24

In step S16, it is determined whether the user reads the data item. The screen control signals from the display control device 33 operated by the user are sent to the server 10, where the user profile generator 24

5   determines from the screen control signal the speed at which the user is scrolling through the data item. The user profile generator 24 also analyses the data signals indicating the momentary contents of the display screen to find the number of words in the text currently

10   displayed on the screen, and then refers to the look-up table 24a to find the screen scrolling speeds which indicate attentive reading and speed reading, respectively, of the displayed text. By comparing the scrolling speed indicated by the control signals with the

15   scrolling speeds given in the look-up table 24a, the user profile generator 24 determines whether the data item has been attentively read or skimmed.

For non-textual data items, such as pictures, sounds etc., the user profile generator 24 analyses the data

20   signals indicating the momentary contents of the display screen to find characteristics of the non-textual data item such as picture size, picture content complexity, size of sound file etc. The user profile generator then refers to the look-up table 24a to find the screen

25   scrolling speeds which indicate attentive or speed reading/viewing/listening/experiencing, respectively of the non-textual data item (entries not shown in Figures).

25

The values in the look-up table 24a may be a standard set of scrolling speed ranges. Alternatively, profile generator 24 may measure the user's normal reading speed, for example by asking the user to read a test text during a sign-up session and measuring the time taken. The profile generator may then compile a look-up table 24a based on the number of words in the test text and the user's measured reading speed.

If it is determined that the user has read the item attentively, the process flows to step S17 where the user profile generator 24 compares the content descriptors of the selected data item with the user's profile to see if they are already included in the profile.

If it is determined in step S16 that the user has not read the data item attentively, the process passes to step S18 where it is determined whether the user has speed-read or 'skimmed through' the item. If the user has scrolled through the data item at a higher speed than is consistent with reading the item attentively (i.e. the scrolling speed is in the second range given by look-up table 24a), it is determined that he has speed-read or 'skimmed through' the item and the process flows to step S17.

If it is determined in step S18 that the user has not speed-read or 'skimmed through' the item, because the user has given an 'exit data item' signal to the processor 23, then the process returns to step S12 and

26

the index and 'of interest' lists are again displayed for the user to select another item.

In step 17, the user profile generator 24 compares the content descriptors CD of the selected data item D
5 with the content descriptors listed in the user's profile. If in step S17 it is determined that the content descriptors of the data item are already included in the user's profile, then the process returns to step S12. If, however, the content descriptors of the data
10 item include elements which are not already in the user's profile, then the process passes to step S19 and the user's profile data is updated to include the new content descriptors. The process then returns to step S5 to compare the data items in the memory 21 with the updated
15 user's profile and display updated index and 'of interest' lists to the user for further selection of data items.

If the user has not selected a data item in step S14 after a determined time interval, the process passes to
20 step S20 and the user is asked if he wishes to exit. If not, he is returned to step S14 to select a data item. If exit is selected, the process ends.

Various alternative implementations are envisaged for the user profile generator 24. The monitoring of the
25 user's screen controls may be effected by relaying the control signals input by the operator to scroll the screen display directly to the profile generator 24 in

27

the server 10 via the distribution network 20.
Alternatively, the user's screen control signals may be
stored temporarily in the user station RAM 39 together
with the data signals indicating corresponding screen

5   contents, and sent to the profile generator 24 when the
user stops perusing a data item. This may be achieved
by uploading this data from the RAM 39 to the profile
generator 24 either when the user exits from a data item
or when the next data item is selected from the index

10   list. In an alternative embodiment, the profile
generator 24 may form part of the user station 30 and may
communicate with the processor 23 at the server station
10 via the distribution network 20. It is further
envisaged that the server station 10 may comprise only

15   the memory 21 and transmitter/receiver, and the user
station may comprise the content analysis unit 22, the
matching unit 27 and the profile generating unit 24. The
user profile register 25 may be at the server or at a
third location, as may the data analysis unit 26. The

20   user profile register 25 is preferably accessible by the
server to obtain user profiles as users access the system
to retrieve data. The operator of the system will also
require the user profile register to be accessible by the
analysis unit 26 to conduct statistical analysis of the

25   users' profiles and possibly correlate them with user
identity details such as address, age, gender, marital
status etc.

28

In yet a further alternative, it is envisaged that the server station 10 may comprise only the memory 21 and transmitter/receiver, and the user station may comprise only a display, a display controller and a

5   transmitter/receiver, and the other elements of the system may be placed at a third location communicating with the server and user stations via the distribution network.

In the system described above, each data item D1,

10 D2 etc is assigned a single content descriptor CD1, CD2... relating to the data item as a whole. In an advantageous embodiment of the system, each data item D1 is subdivided into a number of data item portions D1a, D1b, D1c etc, and each portion is allocated an individual

15 content descriptor CD1a, CD1b, Cd1c... By monitoring the user's screen controls to determine the attention given to each data item portion D1a, D1b etc the content descriptors CD1a, CD1b etc associated with each data item portion may be included or not in the user profile data.

20 Such an arrangement of the data in the memory 21 enables the content descriptors added to the user's profile data more closely to follow the user's interests, since in a large data item the parts through which the user merely speed-reads or skims may be given less weight than those

25 parts which are attentively read when compiling the user profile data.

To implement such an embodiment, the system may be

29

arranged so that when a data item is selected, the server station 10 will download the entire data item to the user station 30, and then break the communication link while the user peruses the data item. A user profile generator

5   provided at the user station 30 stores the user's profile data, and also serves to monitor the user's navigation of the selected data item to determine which portions are read, which are skimmed, and which are ignored. The profile generator then compares the content descriptors

10   of the parts which were read and the parts which were skimmed and optionally the parts which were not read with the user profile stored in the profile generator, and updates the profile as necessary. At the next log on operation, the updated user profile may be uploaded to

15   the user profile register 25 at the server station 10, and may be used to generate an 'of interest' list as described above for sending to the user.

In a further advantageous embodiment of the method, the user profile generator performs periodical "weeding"

20   operations on the user profile data, to remove from the user profile any content descriptors relative to data items that the user no longer finds interesting. This may be achieved by means of the content descriptors presented in step S17 for comparison with the user

25   profile. For example, each content descriptor in the user profile may have associated therewith date or time information showing the last occasion on which the user

30

read or skimmed through a data item having that content

descriptor, and when a predetermined interval has elapsed

after that date, the content descriptor may either have

its weighting reduced, or may be removed from the user's

5   profile. Conversely, every time a user reads a data item

having that content descriptor, the date or time attached

to the content descriptor in the user profile will be

updated, thus ensuring that subject-matter of ongoing

interest remains in the user's profile.   In an

10   alternative "weeding" strategy, a record may be kept each

time the user accesses the database and data items are

matched with content descriptors in the users profile,

and the weighting of such content descriptors may be

decreased if the user does not read the data item and

15   increased if he does. The date and time when a content

descriptor is added to the profile may be recorded, and

a record may also be kept of the date and time on which

data items matching that content descriptor were

detected, and on which the data items were read.

20   The data items D1, D2 etc stored in mass storage

memory 21 may be organised into a number of categories

dependant on their subject matter.

The user may, when first accessing the system, be

presented with a list of available categories and asked

25   to select those that are of immediate interest.   The

processor 23 would then compare only the data items in

those categories with the user profile stored in register

31

25, achieving a significant saving in processing time, before sending the index and 'of interest' lists to the user.

As has been indicated, the data items stored in the mass storage 21 may relate to any field of interest; they may be news items in an electronic newspaper, or may be advertising materials in the form of "small ads" placed by individuals. A further alternative is the use of the system as an electronic 'catalogue' from which users may inspect and order merchandise to be delivered to the user's location. In one such system, the data items will each relate to an individual product, and the data items may be classified generally in 'clothing', 'gardening', 'sports equipment' sections, as well as having specific descriptors associated with each item. By compiling user profiles in the manner described earlier, an electronic catalogue retailer will be able to direct to the attention of the purchaser those items or categories of items in which the purchaser has evinced interest in the past, and may offer incentives, for example to prospective purchasers who 'browse' particular items in the catalogue several times without placing orders. The accumulated information in the user profile, gathered without effort on the part of the user, can be analysed to enable the retailer accurately to target promotional materials.

The user profiles of multiple users may be

32

aggregated to identify broad categories of users and their collective preferences. This information is then used to target promotions at specific users, taking into account the preferences of other users within the same

5   category.

The aggregated data may also be used to identify new product opportunities. For example, if a family of products priced at £1, £2.50, £4 and £5 is presented in order of price and the aggregated passive feedback from

10   users indicates that a lot of time is spent trying to decide whether to purchase the £2.50 or £4 product, this may be taken as an indication that introducing a new product priced at £3 is appropriate.

33

<u>CLAIMS</u>

1.   A data storage and retrieval system comprising:

      storage means to store a plurality of data items;

      selection means for selecting a data item from the

5   said plurality of data items;

      display means capable of displaying the selected

data item;

      display control means generating screen control

signals to vary the display of the data item;

10   monitoring means to receive the screen control

signals and determine a characteristic of the data

display;

      display analysing means to analyse the momentary

content of the display and establish an expected value

15   of the characteristic of the data display;

      inferring means to compare the expected value of the

characteristic with the actual value of the

characteristic, and infer a level of user interest in the

data item therefrom;

20   means to analyse the content of a data item and

assign content descriptor information to the data item;

      means to identify the content descriptor information

relating to data items for which a level of interest

above a predetermined threshold has been inferred;

25   means to compile a listing of such identified

content descriptor information.

34

2.   A data storage and retrieval system according to
claim 1, wherein the display means, the display control
means and the selection means are situated at one of a
plurality of user stations;

5   the storage means is situated at a server station;
and

           a distribution network connects the user station
with the server station.

10   3.   A data storage and retrieval system according to
claim 2, wherein a display analysing means, a monitoring
means, an inferring means and an identifying means are
situated at each user station.

15   4.   A data storage and retrieval system according to any
preceding claim wherein the data display characteristic
is a vertical screen scrolling rate.

5.   A data storage and retrieval system according to
20   claims 1 to 3, wherein the data display characteristic
is a length of time during which data is displayed.

6.   A data storage and retrieval system according to any
preceding claim, wherein the storage means to store the
25   plurality of data items comprises a number of storage
devices located separately from one another.

35

7.   Data storage and retrieval apparatus, comprising:

mass storage means for storing a plurality of data items each having content descriptor information associated therewith;

5   display means for displaying a data item to a user;

control means operable by the user to provide selection signals to select a data item for display, and to provide screen control signals to the display means to navigate within the displayed data item;

10   inferring means which receives the selection signals and the screen control signals, determines from the selection signals which data item is displayed, and infers from the screen control signals a level of interest in the data item;

15   comparison means to determine whether the level of interest in a data item exceeds a threshold value;

user profile generating means to record the content descriptor information of data items whose level of interest exceeds the threshold value in a user profile

20   memory;

matching means to compare the content descriptor information of the data items in the mass storage memory with the content descriptor information in the user profile memory;

25   means to indicate to the user those data items whose content descriptors match the content descriptors held in the user profile memory.

36

8. Apparatus according to claim 7, wherein the mass storage memory is situated at a server station and wherein a plurality of display means and associated control means are provided at respective user stations,

5   and the user stations and server station are connected via a distribution network providing two-way communication between each user station and the server station.

10 9. Apparatus according to claim 8, wherein the server station further comprises the matching means.

10. Apparatus according to claim 8 or claim 9, wherein the user station further comprises the inferring means,

15 the comparison means and the user profile generating means.

11. Apparatus according to claim 8 or claim 9, wherein the inferring means, the comparison means and the user

20 profile generating means are situated at a third location separate from the server station and the user stations, but connected thereto by the distribution system.

12. A user station for a data storage and retrieval

25 apparatus according to claim 8 comprising a display device, control means to provide selection and screen control signals, and transmitting means to transmit the

37

screen control signals to the inferring means.

13. A user station for an apparatus according to claim 8, wherein the user station further includes an inferring means, a comparison means and a user profile generating means.

14. A user station according to claim 13, further including a user profile memory, and means to transmit the contents of the user profile memory to the matching means.

15. A server station for a data storage and retrieval apparatus according to claim 8, further comprising the inferring means, the comparison means, the user profile generating means and the matching means.

16. Data storage and retrieval apparatus according to claim 7, further including data content analysis means to analyse the content of data items and allocate content descriptors thereto.

17. A method of determining the level of interest that data displayed on a display device has for a reader, comprising displaying the data to the reader on a display device controlled by a controller by means of which the reader inputs control signals to vary the display,

38

wherein a processor receives the input signals and determines therefrom a rate at which the screen display is changed, and the processor compares the rate of change with an expected rate of change to determine the user's

5   interest level.

18.   A method according to claim 17, wherein the expected rate of change is established by determining the number of words of text displayed on the screen and correlating

10   the number of words with a predetermined rate of change corresponding thereto.

19.   A method according to claim 18, wherein the number of words of text displayed is determined by counting.

15

20.   A method according to claim 18, wherein the number of words of text displayed is determined by calculation based on the text area and the font size.

20   21.   A method of discriminating between data of interest to a user and other data, comprising the steps of storing in a memory information relating to the interests of the user;

     presenting a number of data items to the user for

25   selection;

     displaying a selected data item on a display controllable by the user for navigation within the data

39

item;

monitoring the display control signals input by the user during navigation and determining therefrom the user's level of interest in the data item displayed,

5   determining whether the level of interest exceeds a threshold and adding information relating to the content of data items whose level of interest exceeds the threshold to a user profile;

comparing information relating to the content of the

10   data items with the information stored in the user profile;

determining those data items whose content matches the information in the user profile to be of interest to the user;

15   modifying the presentation of the data items so as to present data items determined to be of interest in a form distinct from other data items.

22. A method according to claim 21, wherein the step of

20   presentation of the data items for selection comprises displaying a listing of the data items on a display device controllable by the user to input a selection signal to select a data item from the displayed listing.

25   23. A method according to claim 21 or claim 22, wherein the step of determining the user's interest level comprises analysing the content of the display to

40

establish the amount of text being displayed to the user,
calculating from the display control signals the speed
with which the user is moving through the text, comparing
the amount of text displayed with the speed of movement

5   to deduce a reading speed, and comparing the reading
speed so deduced with predetermined reading speed ranges
corresponding to different levels of interest to infer
the user's level of interest in the displayed data item.

10   24.   A method according to claim 23, wherein the analysis
of the content of the display comprises a word count of
the displayed text.

25.   A method according to claim 23, wherein the display

15   content is analysed to measure the area of the display
devoted to text and the font size of the text, and from
these measurements a determination of the number of words
displayed is made.

20   26.   A method according to claims 21 or 22, wherein the
step of determining the user's interest level comprises
analysing the display content to determine the type
and/or amount of non-textual data items being displayed
to the user, calculating from the display control signals

25   the speed with which the user is moving through a non-
textual data item, comparing the type and/or amount of
the non-textual data item to deduce a reading/viewing/

experiencing speed, and comparing the reading/viewing/
experiencing speed so deduced with predetermined reading/
viewing/experiencing speed ranges corresponding to
different levels of interest to infer user's level of
5   interest in the displayed data item.

27.   Data storage means carrying written instructions for
causing a processing device to perform the steps of:
storing in a memory information relating to the
10   interests of the user;
presenting a number of data items to the user for
selection;
displaying a selected data item on a display
controllable by the user for navigation within the data
15   item;
monitoring the display control signals input by the
user during navigation and determining therefrom the
user's level of interest in the data item displayed,
determining whether the level of interest exceeds
20   a threshold and adding information relating to the
content of data items whose level of interest exceeds the
threshold to a user profile;
comparing the content of the data items with the
information stored in the user profile;
25   determining those data items whose content matches
the information in the user profile to be of interest to
the user;

42

modifying the presentation of the data items so as to present data items determined to be of interest in a form distinct from other data items.

5   28. A data storage medium carrying processor-implementable instructions for causing a processing apparatus to perform a method of any of claims 17 to 26.

29. An   electrical   signal   carrying   processor

10   implementable instructions for causing a processing apparatus to perform a method of any of claims 17 to 26.

30. Monitoring apparatus for a terminal having a display device and user-operable input means for inputting

15   display commands to control a characteristic of the display, the monitoring apparatus comprising:

means for identifying the content of data displayed on the display device;

means for correlating display commands input by a

20   user with the content of displayed data; and

means for outputting a signal representative of the correlation between the content of data displayed and the display commands input by the user.

25   31. Monitoring apparatus according to claim 30 wherein the monitoring apparatus forms part of the terminal.

43

32.   Monitoring apparatus according to claim 30, wherein
the monitoring apparatus is situated remote from the
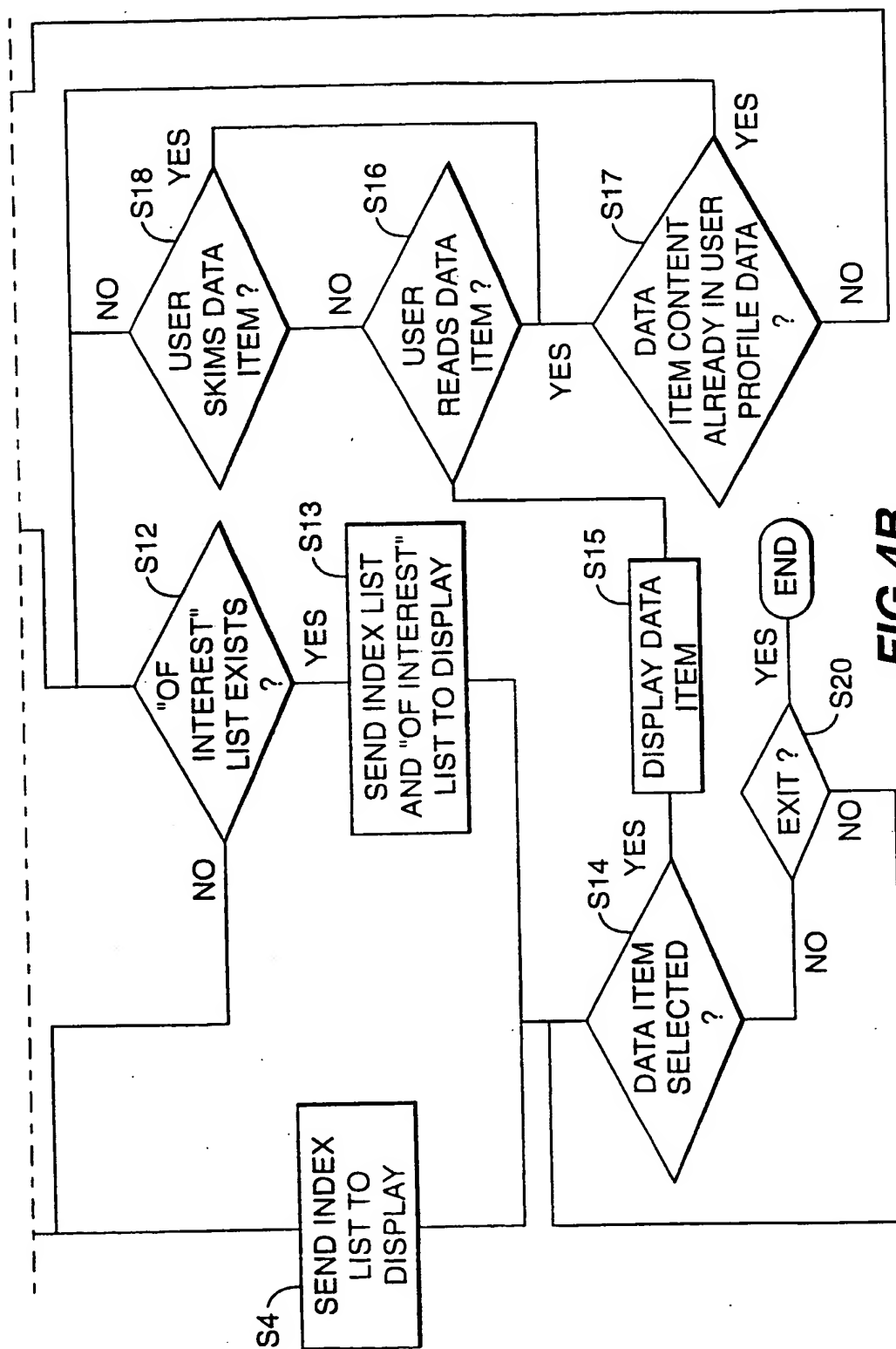terminal and receives signals therefrom.

5

FIG. 1
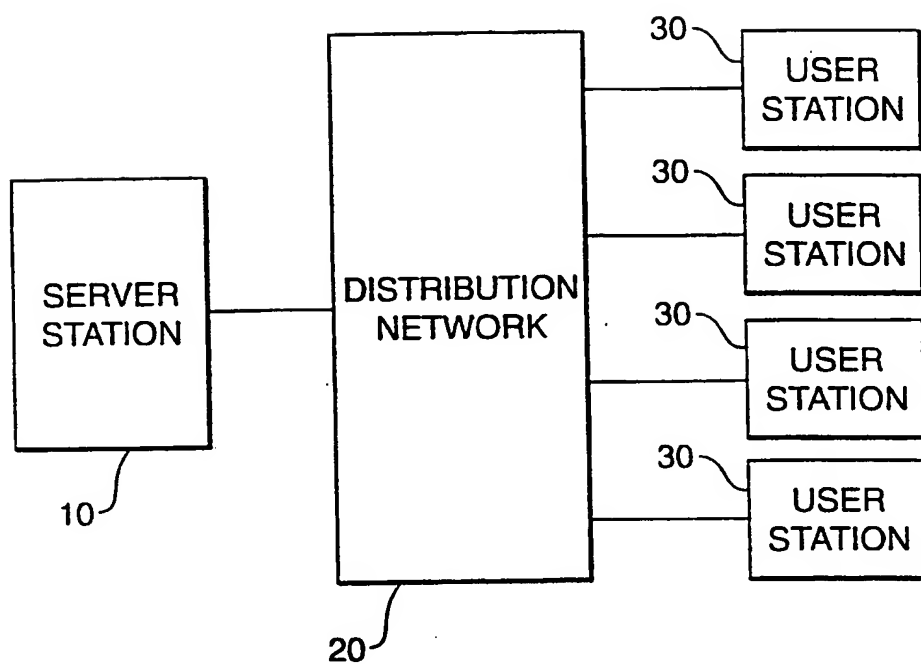
FIG. 2

FIG. 3

FIG.4

FIG.4A

FIG.4B

FIG.4A
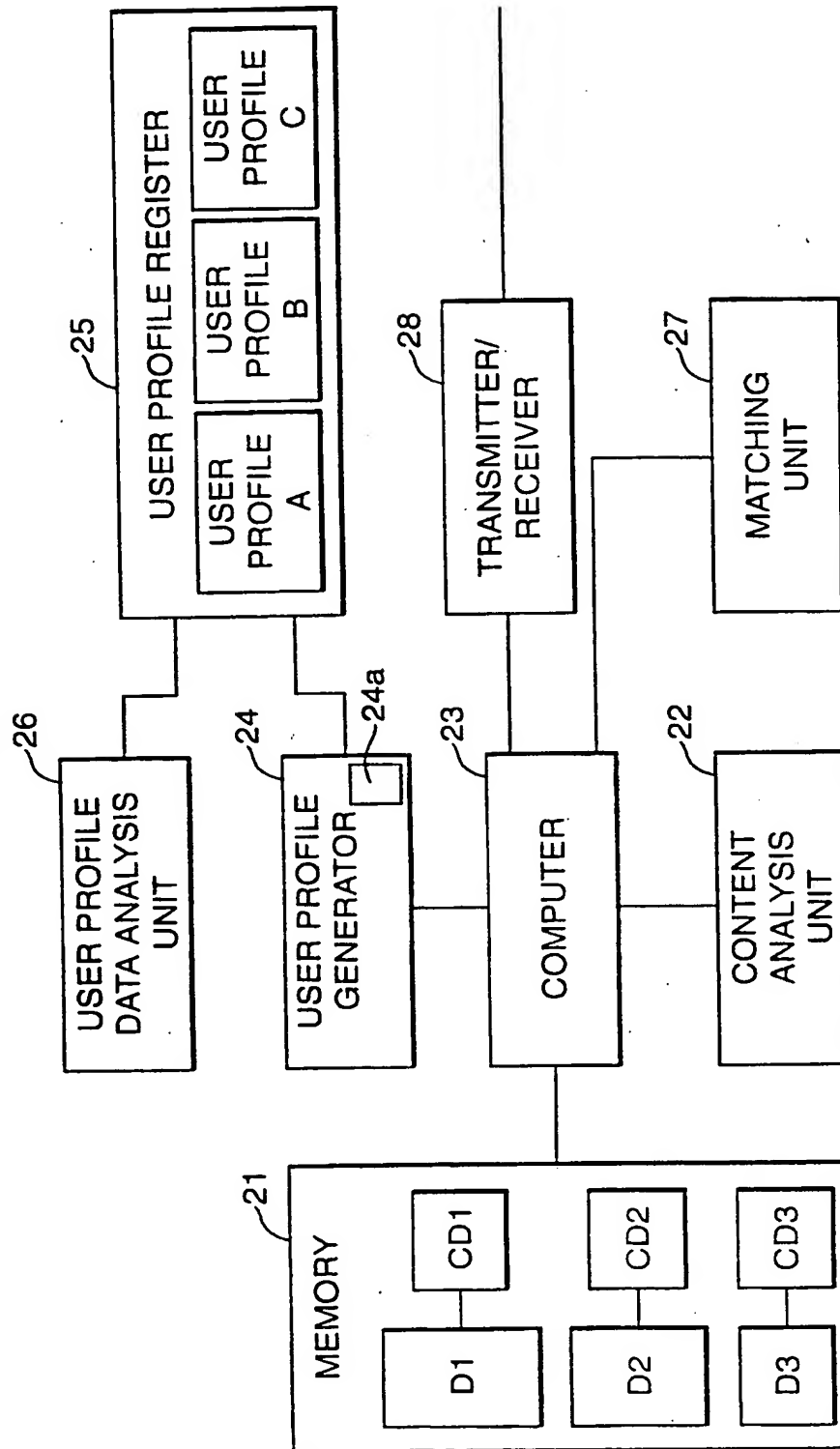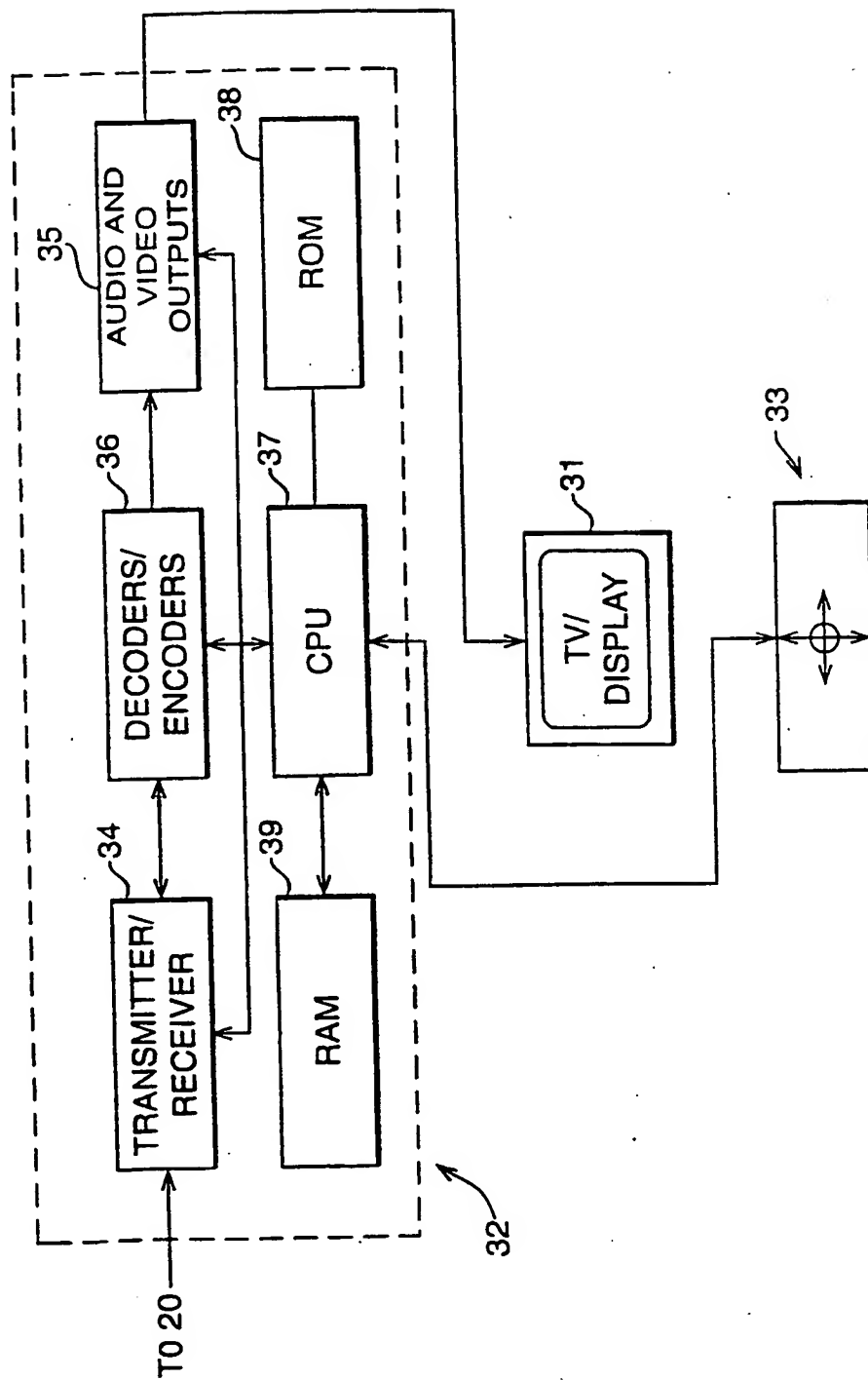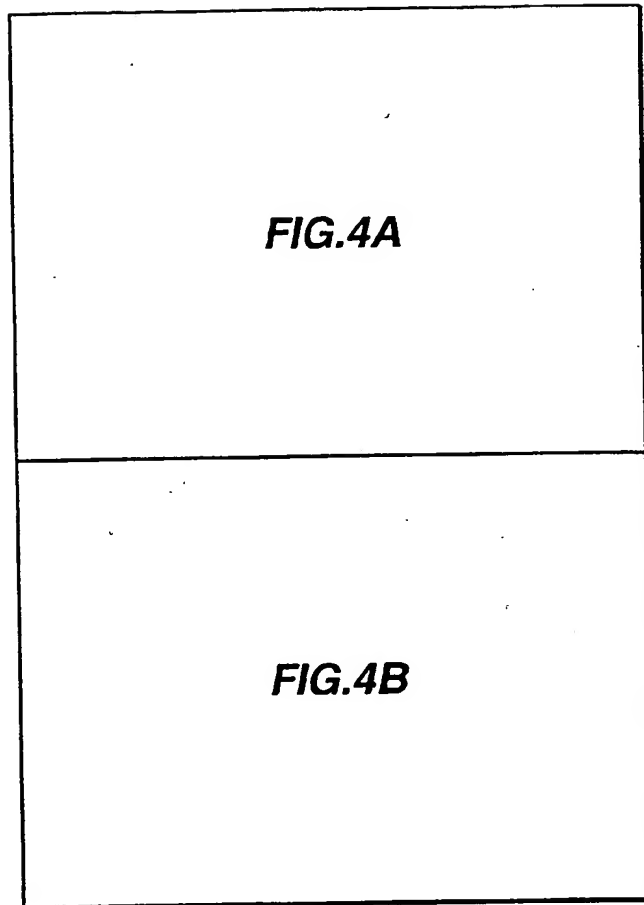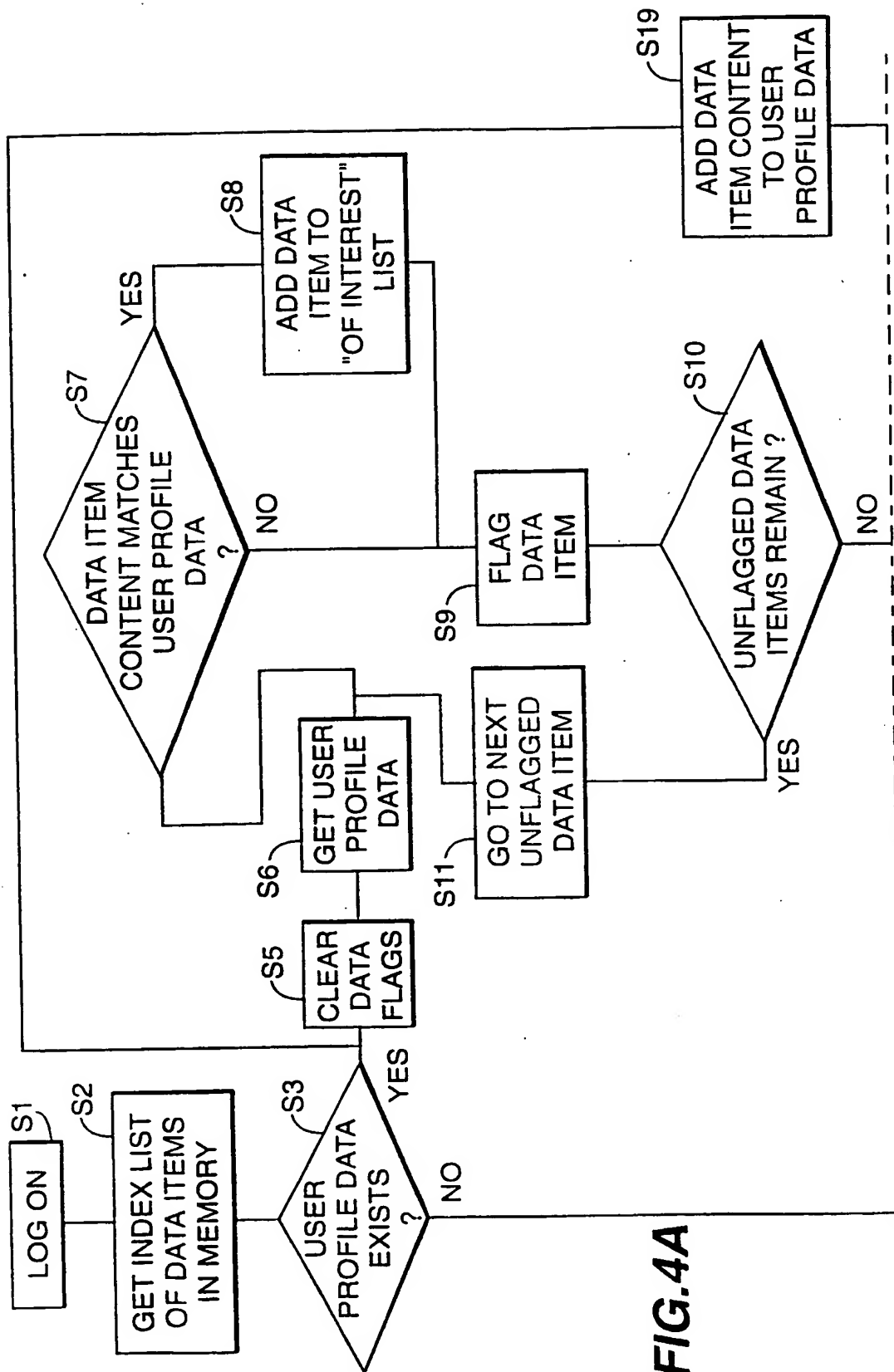
FIG.4B

FIG. 5

No. OF WORDS ON SCREEN	SCROLL SPEED FOR ATTENTIVE READING	SCROLL SPEED FOR SPEED READING
20	2-3 SECONDS/SCREEN	1-2 SECONDS/SCREEN
50	8-10 SECONDS/SCREEN	2-3 SECONDS/SCREEN
100	16-20 SECONDS/SCREEN	4-6 SECONDS/SCREEN
150	24-30 SECONDS/SCREEN	8-12 SECONDS/SCREEN
200	32-40 SECONDS/SCREEN	10-14 SECONDS/SCREEN

24a

FIG. 6
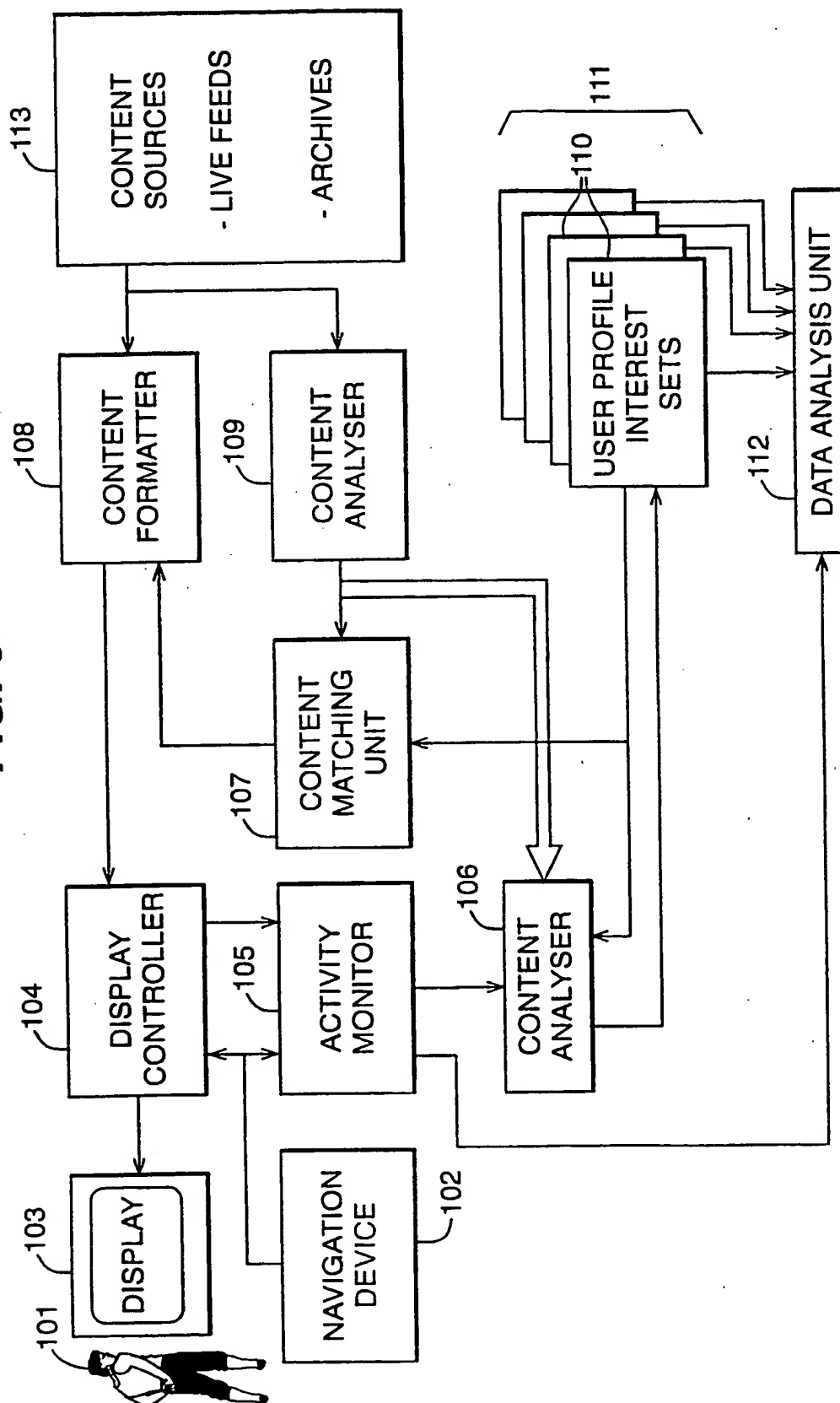
INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6   G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED
Minimum documentation searched (classification system followed by classification symbols)
IPC 6   G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KAMBA T ET AL: "ANATAGONOMY: a personalized newspaper on the World Wide Web" INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES, JUNE 1997, ACADEMIC PRESS, UK, vol. 46, no. 6, pages 789-803, XP002086827 ISSN 1071-5819 see the whole document -/--	1-32

[X] Further documents are listed in the continuation of box C.		[X] Patent family members are listed in annex.

Date of the actual completion of the international search	Date of mailing of the international search report
4 December 1998	21/12/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Abbing, R

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 446 891 A (KAPLAN CRAIG A ET AL) 29 August 1995 see abstract see column 11, line 28 – column 16, line 64 see claims ——	1-3, 5-10, 13-19, 21-24, 27-31

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5446891 A	29-08-1995	NONE	